

SKRIPSI

**PENGUNAAN *PERSON-FIT* PADA ITEM *POLYTOMOUS*  
DAN HUBUNGANNYA DENGAN JENIS KELAMIN DAN  
ETNIK**



**Disusun oleh:**

**Maria Mona Holaw**

**155120301111011**

**PROGRAM STUDI S1 PSIKOLOGI**

**FAKULTAS ILMU SOSIAL DAN ILMU POLITIK**

**UNIVERSITAS BRAWIJAYA**

**2018**

## LEMBAR PENGESAHAN


**Penggunaan *Person-Fit* pada Item *Polytomous* dan Hubungannya dengan  
Jenis Kelamin dan Etnik**

### SKRIPSI

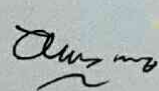
Disusun Oleh :  
**Maria Mona Holaw**  
**NIM. 155120301111011**

Telah disetujui dan dinyatakan lulus dalam ujian sarjana  
Pada tanggal **25 November 2018**

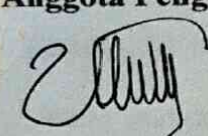
**Tim Penguji**  
**Ketua Majelis Sidang Penguji,**

  
Sukaesi Marianti, S.Psi., M.Si., Ph.D.  
NIP. 198007312008122002

**Ketua Penguji**

  
Nur Hasanah, S.Psi., M.Si.  
NIP. 197404022008012012

**Anggota Penguji,**

  
Ulifa Rahma, S.Psi., M.Psi, Psikolog  
NIK. 2013098803132001

**03 DEC 2018**  
Malang,  
Fakultas Ilmu Sosial dan Ilmu Politik

**Dekan**



  
Prof. Dr. Unti Ludigdo, Ak  
NIP. 19690814 199402 1 001

## LEMBAR PERNYATAAN ORISINALITAS

Nama: Maria Mona Holaw

NIM: 155120301111011

Menayatakan dengan kesungguhan bahwa skripsi dengan judul **“Penggunaan *Person-Fit* pada Item *Polytomous* dan Hubungannya dengan Jenis Kelamin dan Etnik”** adalah benar-benar karya saya sendiri. Hal yang bukan karya saya dalam skripsi tersebut diberi tanda kutipan dan ditunjukkan dalam daftar pustaka.

Apabila di kemudian hari terbukti bahwa pernyataan saya tidak benar, maka saya bersedia menerima sanksi akademik berupa pencabutan skripsi dan gelar yang saya peroleh dari skripsi tersebut.

Malang, 28 November 2018

Yang membuat pernyataan



Maria Mona Holaw

155120301111011



## KATA PENGANTAR

Puji syukur penulis panajtkan kepada Allah SWT karena dengan kuasa-Nya penulis dapat menyelesaikan skripsi yang berjudul “Penggunaan *Person-Fit* pada Item *Polytomous* dan Hubungannya dengan Jenis Kelamin dan Etnik”. Skripsi ini merupakan tugas akhir guna memenuhi persyaratan kelulusan pada Jurusan Psikologi Fakultas Ilmu Politik dan Ilmu Budaya.

Penulis menyadari dalam menyelesaikan skripsi ini tidaklah mudah, dan membutuhkan bantuan serta dukungan dari berbagai pihak. Oleh karena itu, penulis inhgin mengucapkan terimakasih kepada:

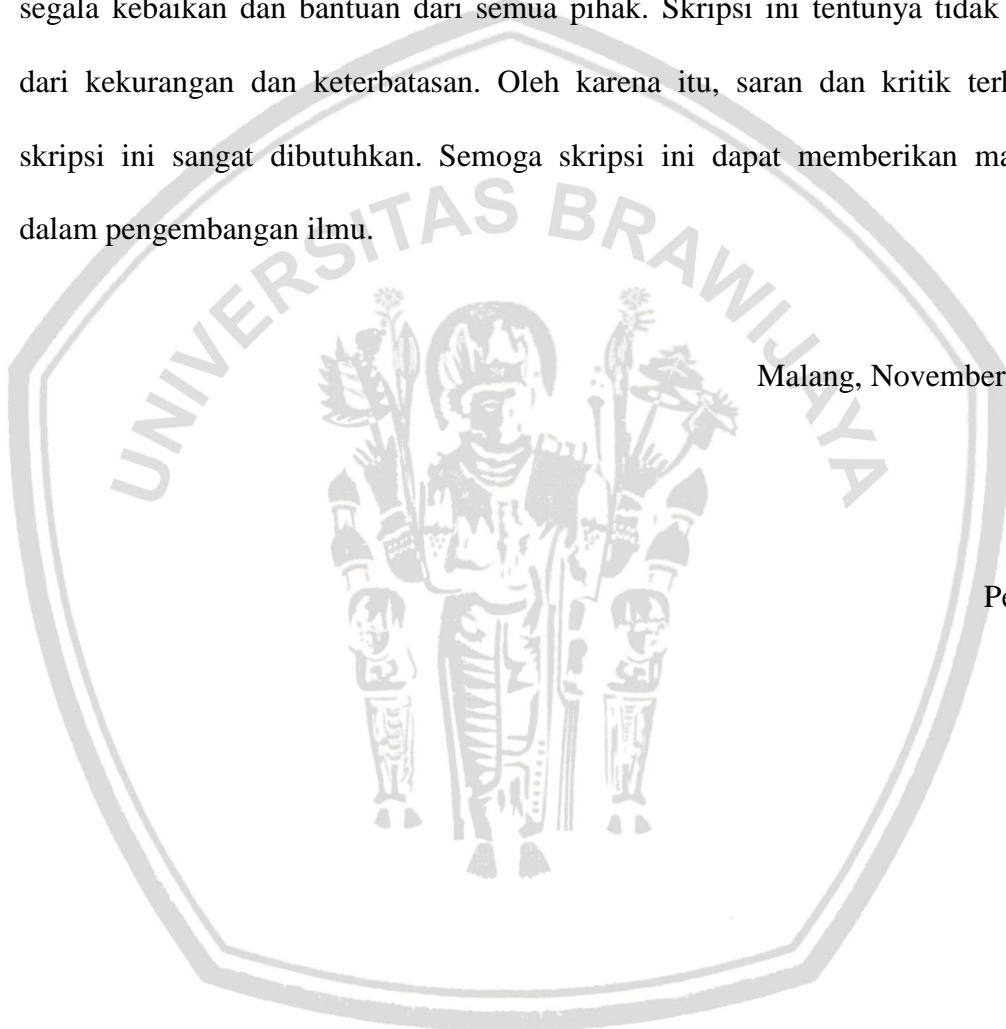
1. Orang tua dan keluarga yang telah mendukung dan memberikan semangat selama pengerjaan skripsi
2. Ibu Sukaesi Marianti, Ph.D, selaku dosen pembimbing yang dengan sabar mengarahkan, memberikan bimbingan dan semangat bagi penulis dalam menyelesaikan skripsi.
3. Ibu Nur Hasanah, S.Psi., M.Psi, dan Ibu Ulifa Rahma S.Psi., M.Psi., Psikolog, selaku dosen penguji yang telah membantu menyempurnakan skripsi ini.
4. Sahabat penulis, Agustina Rosianawati dan Rekyan Puspitasari yang senantiasa mendukung dan menyemangati penulis.
5. Team laboratorium yang membantu penulis dalam menyelesaikan skripsi.
6. Seluruh rekan mahasiswa Psikologi angkatan 2015 yang telah banyak membantu dan mendukung penulis.

7. Seluruh pihak yang telah menyemangati dan memotivasi penulis dalam menyelesaikan skripsi.

Akhir kata, penulis bereharap Tuhan Yang Maha Esa dapat membalas segala kebaikan dan bantuan dari semua pihak. Skripsi ini tentunya tidak lepas dari kekurangan dan keterbatasan. Oleh karena itu, saran dan kritik terhadap skripsi ini sangat dibutuhkan. Semoga skripsi ini dapat memberikan manfaat dalam pengembangan ilmu.

Malang, November 2018

Penulis



## DAFTAR ISI

LEMBAR PENGESAHAN .....	i
LEMBAR PERNYATAAN ORISINALITAS .....	ii
KATA PENGANTAR.....	iii
DAFTAR ISI.....	v
DAFTAR TABEL .....	viii
DAFTAR GAMBAR.....	ix
DAFTAR LAMPIRAN .....	x
DAFTAR ISTILAH .....	xi
ABSTRAK .....	xii
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
A. Latar Belakang .....	1
B. Rumusan Masalah .....	5
C. Tujuan Penelitian.....	6
D. Manfaat Penelitian .....	6
E. Penelitian Terdahulu.....	6
<b>BAB II KAJIAN TEORI .....</b>	<b>9</b>
A. <i>Person-Fit</i> .....	9
1. Pendekatan <i>Person-Fit</i> .....	9
2. Statistik <i>Person-Fit</i> .....	10
3. Pola Respon <i>Aberrant</i> .....	12
B. Model <i>Polytomous</i> dalam IRT .....	13

C. Korelasi <i>Person-Fit</i> dengan Gender dan Etnik .....	15
D. Identitas Etnik .....	17
1. Definisi Identitas Etnik.....	17
2. Dimensi Identitas Etnik .....	18
3. Skala Identitas Etnik.....	19
E. Variabel Penelitian. ....	19
F. Kerangka Pemikiran .....	20
<b>BAB III METODE PENELITIAN .....</b>	<b>21</b>
A. Desain Penelitian.....	21
B. Fokus Penelitian .....	21
C. Definisi Operasional.....	22
D. Subjek Penelitian.....	22
E. Tahap Penelitian .....	23
F. Teknik Analisis Data .....	24
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>25</b>
A. Gambaran Umum Subjek penelitian .....	25
1. Parameter Item dengan <i>Graded Response Model</i> .....	26
2. Skor <i>Person-fit</i> Statistik $I_z$ .....	28
3. Uji Beda <i>Person-fit</i> berdasarkan Jenis Kelamin dan Etnik.....	30
B. Hasil Penelitian .....	26
C. Pembahasan.....	32
D. Keterbatasan Penelitian .....	38
<b>BAB V PENUTUP .....</b>	<b>39</b>

A. Kesimpulan .....	39
B. Saran.....	40
<b>DAFTAR PUSTAKA .....</b>	<b>41</b>





## DAFTAR TABEL

Tabel 1. Variabel penelitian .....	19
Tabel 2. Kategorisasi subjek penelitian .....	25
Tabel 3. Parameter item dengan <i>Graded Response Model</i> .....	27
Tabel 4. $l_z$ value yang terdeteksi sebagai <i>aberrance</i> .....	29
Tabel 5. <i>Latent trait</i> peserta tes yang terdeteksi sebagai <i>aberrance</i> .....	30
Tabel 6. Korelasi $l_z$ value dengan jenis kelamin dan etnik peserta tes.....	30
Tabel 7. Perbandingan <i>mean</i> $l_z$ value berdasarkan jenis kelamin dan etnik.....	31



## DAFTAR GAMBAR

Gambar 1. Kerangka Penelitian .....	20
Gambar 2. Desain penelitian <i>person-fit</i> pada item <i>polytomous</i> .....	21
Gambar 3. $I_z$ value peserta tes .....	28
Gambar 4. Pola respon peserta tes yang terdeteksi sebagai <i>aberrance</i> .....	29



## DAFTAR LAMPIRAN

Lampiran 1. Blueprint .....	44
Lampiran 2. Skala identitas etnik.....	45
Lampiran 3. <i>Output Graded Respose Model</i> .....	47
Lampiran 4. <i>Item Characteristic Curves</i> .....	48
Lampiran 5. <i>Item Operation Characteristic Curves</i> .....	54
Lampiran 6. <i>Output person-fit</i> .....	58
Lampiran 7. <i>Latent Trait</i> .....	64
Lampiran 8. <i>Output Multivariate Anova</i> .....	69
Lampiran 9. Surat permohonan ijin penelitian skripsi .....	70
Lampiran 10. Berita acara seminar proposal.....	71
Lampiran 11. Nilai bimbingan skripsi .....	72
Lampiran 12. Kartu kendali skripsi.....	73

## DAFTAR ISTILAH

**Aberrance** Peserta tes yang terdeteksi memberikan pola respon yang dianggap *aberrant*.

**Aberrant** Pola respon yang menyimpang atau atipikal dari peserta tes.

**Blueprint** Tabel yang berisi informasi mengenai proporsi item pada skala.

**EIS (*Ethnic Identity Scale*)** Skala identitas etnik yang dibuat oleh Umaña-Taylor, Yazedjian, dan Bámaca-Gómez (2004) dan telah ditransadaptasi ke bahasa Indonesia.

**Item Discrimination** Tingkatan suatu item dalam membedakan peserta tes dengan latent trait tinggi dan rendah.

**IRT (*Item Response Theory*)** Teori analisis modern guna mengukur suatu skala berdasarkan hubungan antara probabilitas respon, item pada tes, dan kemampuan atau *latent trait* peserta tes.

**Latent Trait** Nilai sifat yang diukur dari peserta tes.

**$I_z$  value** Skor  $I_z$  yang mengindikasikan nilai *person-fit* tiap peserta tes.

**Non-aberrance** Peserta tes yang terdeteksi memiliki pola respon yang wajar atau tipikal.

**Person-fit** Metode statistik yang digunakan untuk mendeteksi pola respon menyimpang pada suatu data set.

**Statistik  $I_z$**  Salah satu statistik *parametric* dalam *person-fit* yang didasari oleh statistik  $l_o$  yang menggunakan maksimum fungsi *likelihood*.



## Penggunaan *Person-Fit* pada Item *Polytomous* dan Hubungannya dengan Jenis Kelamin dan Etnik

Maria Mona Holaw  
155120301111011

### ABSTRAK

Penelitian ini bertujuan untuk mengetahui kemampuan statistik  $l_z$  dalam mendeteksi pola respon yang menyimpang dari *data set*, dan untuk mengetahui hubungan *person-fit* dengan jenis kelamin dan etnik peserta tes. Penelitian ini menggunakan *Ethnic Identity Scale* (EIS) sebagai data *polytomous* yang akan dideteksi dengan *person-fit*. Subjek penelitian sebagai peserta tes merupakan 560 siswa SMA Kolese Santo Yusup Kota Malang. *Software* yang digunakan adalah R Studio versi 3.2.2 untuk melakukan analisis data. Hasil penelitian ini menemukan bahwa  $l_z$  mendeteksi sebanyak 8 peserta tes (1.42%) sebagai *aberrance*. Penelitian ini juga menemukan bahwa terdapat perbedaan *person-fit* yang signifikan berdasarkan etnik peserta tes.

Kata kunci: *aberrant*, *aberrance*, *person-fi*, statistik  $l_z$ , jenis kelamin, etnik.

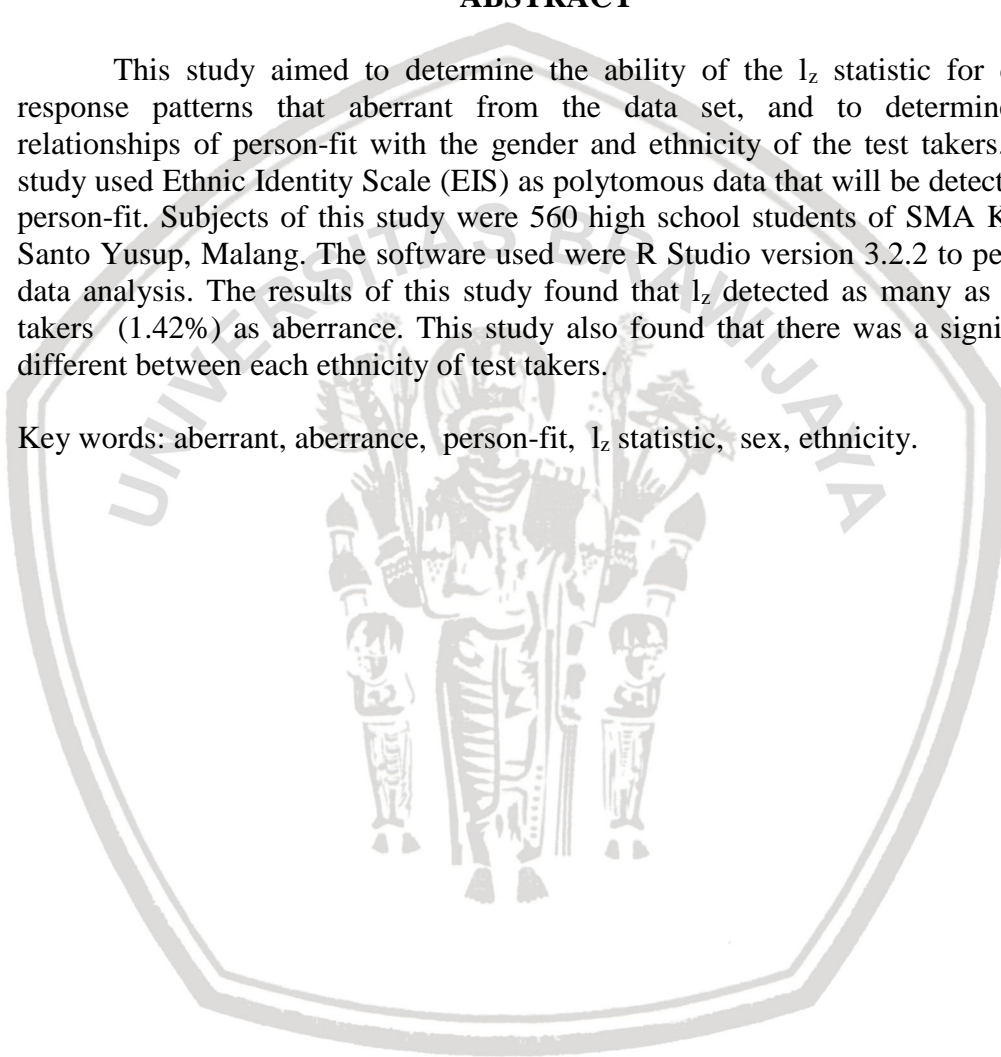
## The Application of Person-Fit on Polytomous Items and its Relationship to Sex and Ethnicity

Maria Mona Holaw  
155120301111011

### ABSTRACT

This study aimed to determine the ability of the  $I_z$  statistic for detect response patterns that aberrant from the data set, and to determine the relationships of person-fit with the gender and ethnicity of the test takers. This study used Ethnic Identity Scale (EIS) as polytomous data that will be detected by person-fit. Subjects of this study were 560 high school students of SMA Kolese Santo Yusup, Malang. The software used were R Studio version 3.2.2 to perform data analysis. The results of this study found that  $I_z$  detected as many as 8 test takers (1.42%) as aberrance. This study also found that there was a significant different between each ethnicity of test takers.

Key words: aberrant, aberrance, person-fit,  $I_z$  statistic, sex, ethnicity.



## BAB I

### PENDAHULUAN

#### A. Latar Belakang

Pengukuran psikologis menggunakan alat tes akan menghasilkan informasi mengenai variabel yang diukur. Informasi tersebut kemudian dapat digunakan untuk berbagai kepentingan, misalnya untuk mengidentifikasi karakteristik individu, mengurutkan individu berdasarkan variabel yang diukur, hingga menyeleksi individu berdasarkan karakteristik tertentu (Karabatsos, 2003). Oleh sebab itu, pengukuran menggunakan alat tes memerlukan akurasi yang baik.

Menurut Bracey (dalam Meijer, 1996) studi mengenai akurasi lebih difokuskan pada karakteristik tes yang mempengaruhi akurasi hasil tes. Studi mengenai akurasi menemukan bahwa terdapat beberapa faktor yang mempengaruhi tingkat keakuratan dari pengukuran menggunakan alat tes, misalnya karakteristik administrasi tes (instruksi yang kurang jelas, kesalahan skoring), dan karakteristik item (penggunaan bahasa yang salah, item yang ambigu) (Meijer, 1996). Selanjutnya dijelaskan bahwa belum banyak studi yang berfokus pada karakteristik peserta tes dalam mempengaruhi akurasi (Meijer, 1996).

Karakteristik peserta tes yang mempengaruhi akurasi berkaitan erat dengan pola respon yang menyimpang. Adanya kepentingan akurasi tes tersebut melandasi dilakukannya studi untuk mendeteksi pola respon item yang

menyimpang (*aberrant*) sekitar tahun 1981 (Meijer, 1996). Pola respon item yang menyimpang (*aberrant*) dapat disebabkan oleh berbagai hal, misalnya *cheaters*, *lucky guessers*, *fumblers*, *plodders*, *language deficiency*, *random guessers or constant responders*, *creatives*, dan *palm beachers* (Meijer, 1996).

*Cheaters* merujuk pada peserta tes yang melakukan kecurangan sehingga dapat memperoleh skor yang tinggi. *Lucky guessers* merujuk pada peserta tes yang menebak-nebak jawaban dengan skor tinggi pada tes. *Flumbers* merujuk pada peserta tes yang kebingungan pada awal pengerjaan tes sehingga mendapatkan skor yang rendah. *Plodders* merujuk pada peserta tes yang tidak menyelesaikan tes hingga item terakhir. *Language deficiency* merujuk pada peserta tes yang tidak memahami bahasa yang digunakan pada tes. *Random guessers* merujuk pada peserta tes yang menebak-nebak jawaban masing-masing item secara acak. *Creatives* merujuk pada peserta tes yang merespon tanpa memperhatikan konten tiap item. *Palm beachers* merujuk pada peserta tes yang salah dalam mengisi lembar jawaban (Meijer, 1996).

Contoh sederhana pola respon menyimpang adalah peserta tes dapat memanipulasi jawaban atau respon terhadap item yang disesuaikan dengan harapan atau tujuan asesmen. Contoh lain dari pola respon yang menyimpang adalah beberapa orang mungkin memberikan tanggapan atau respon secara acak karena mereka tidak termotivasi untuk mengikuti tes (Glas & Dagohoy, 2007). Kedua contoh tersebut akan membentuk pola respon yang atipikal atau dianggap menyimpang karena tidak menggambarkan *latent trait* peserta tes yang bersangkutan.



Metode untuk menguji keakuratan tes adalah dengan mengevaluasi kecocokan kinerja peserta tes dengan *Item Response Theory* (IRT) yang disebut dengan metode *person-fit* (Meijer & Sijsma, 2001). Metode *person-fit* mengacu pada metode statistik untuk mengevaluasi pola respon peserta terhadap suatu tes (Meijer & Sijsma, 2001).

Terdapat pembagian statistik *person-fit* yang dapat digunakan, salah satu statistik yang banyak digunakan adalah statistik  $I_z$  dengan alasan pendukung. Statistik  $I_z$  merupakan salah satu statistik *person-fit* yang paling populer dalam asesmen. Penelitian terdahulu juga menemukan bahwa statistik  $I_z$  memiliki kinerja lebih baik daripada statistik *person-fit* yang lain di banyak kasus dan statistik  $I_z$  telah direkomendasikan sebagai satu dari statistik *person-fit* yang paling akurat untuk pendeteksian perilaku menyimpang (Drasgow, Levine, & Williams, 1985).

Schmitt, Chan, dan Sacco (1999) melakukan studi dengan menghubungkan *person-fit* dengan beberapa variabel, yaitu tes reaksi, motivasi mengerjakan tes, ketelitian, jenis kelamin, dan ras. Studi tersebut menunjukkan bahwa *person-fit* memiliki hubungan dengan variabel-variabel tersebut. Berdasarkan jenis kelamin, hasil studi menunjukkan bahwa peserta tes laki-laki memiliki nilai statistik *person-fit* negatif lebih tinggi daripada peserta tes perempuan. Hasil tersebut memberi indikasi bahwa pola respon menyimpang lebih banyak ditunjukkan oleh laki-laki. Berdasarkan ras, hasil studi menunjukkan bahwa peserta tes dengan ras Afrika-Amerika memiliki nilai statistik *person-fit* negatif yang lebih tinggi daripada peserta tes dengan

ras kulit putih. Hasil tersebut menunjukkan indikasi bahwa pola respon menyimpang lebih banyak ditunjukkan oleh peserta tes dengan ras Afrika-Amerika.

Meijer dan Sijtsma (2001) memberikan gambaran tentang *person fit* yang diusulkan untuk berbagai model IRT. Kebanyakan statistik *person fit* dikembangkan untuk model IRT untuk item *dichotomous*, atau item dengan dua kategori respon (Meijer & Sijtsma, 2001), contohnya penelitian oleh (Drasgow, Levine, & McLaughlin, 1991; Drasgow, Levine, & Williams, 1985; Meijer & Sijtsma, 2001; Smith, 1986; Tatsuoka, 1984; Wright & Stone, 1979). *Person fit* untuk item *polytomous* jauh lebih sedikit (Drasgow, Levine, & Williams, 1985), contohnya penelitian oleh (Drasgow et al., 1985; Emons, 2008; Glas & Dagohoy, 2007; van Krimpen-Stoop & Meijer, 2002; Wright & Masters, 1982). Tes dengan item *polytomous* menarik untuk dibahas karena alat ukur psikologis, utamanya untuk variabel non kognitif didominasi oleh model item *polytomous*. Misalnya, *Self-esteem Scale*, *Locus of Control Scale*, *Self-efficacy Scale*, dan *Ethnic Identity Scale*.

Beberapa tes *polytomous* masuk ke dalam golongan *low stake test*. *Low stake test* merupakan tes yang evaluasinya tidak berdampak besar atau tidak memiliki konsekuensi pribadi bagi peserta tes. Salah satu contoh skala yang termasuk low stake adalah skala *Ethnic Identity Scale* (EIS). Secara sederhana, EIS mengukur identitas etnik seseorang yang hasilnya menunjukkan tinggi atau rendahnya identitas etnik peserta tes tersebut. Hasil tinggi atau rendahnya identitas etnik tersebut tidak akan memiliki dampak tertentu bagi peserta tes.

*Low stake test* telah diasumsikan dan dalam beberapa kasus terbukti terkait dengan penurunan dalam motivasi dan kinerja saat mengerjakan tes (Wise & DeMars, 2005). Kemungkinan motivasi rendah memungkinkan data yang dikumpulkan dari tes kurang valid karena peserta tes dikhawatirkan tidak memberikan respon yang sesuai dengan kemampuan (*ability*) mereka (Wainer, 1993). Berbagai perbedaan jenis kelamin yang terkait dengan motivasi telah ditunjukkan pada *low stake test*. DeMars, Bashkov, dan Socha (2013) menemukan bahwa laki-laki menunjukkan motivasi yang rendah dibanding perempuan dalam mengerjakan tes.

*Ethnic Identity Scale* (EIS) dikembangkan berdasarkan tiga dimensi, yaitu eksplorasi, komitmen, dan afirmasi (Umaña-Taylor, Yazedjian, & Bámaca-Gómez., 2004). EIS banyak digunakan dalam penelitian mengenai identitas etnik, salah satunya adalah penelitian oleh Paramita (2014) yang melakukan transadaptasi skala EIS agar dapat digunakan sesuai budaya Indonesia. EIS terdiri dari 17 item dengan menggunakan model skala Likert yang terdiri dari 4 poin respon. Selanjutnya, skala EIS yang ditransadaptasi Paramita (2014) memiliki nilai reliabilitas *Cronbach Alpha* untuk masing-masing dimensi sebesar .84 pada dimensi eksplorasi, .76 pada dimensi komitmen, dan .88 pada dimensi afirmasi. Lebih lanjut, skala EIS kemudian diuji dengan pendekatan *delayed alternate form* oleh Ningsih (2017). Hasil penelitian Ningsih (2017) menunjukkan hubungan antar dimensi ( $r$ ) terletak diantara .69 – .76. Selanjutnya, teknik hubungan item total ( $rit$ ) menunjukkan bahwa seluruh item memiliki  $rit \geq .30$ .

Secara psikometrik EIS sudah banyak digunakan dan sudah teruji reliabilitasnya di Indonesia. Kendati demikian, penting untuk memperhatikan karakteristik peserta tes dalam mengerjakan skala EIS mengingat EIS termasuk *low stake test*. Penelitian ini bertujuan untuk mengetahui kemampuan *person-fit* dengan statistik  $I_z$  dalam mendeteksi pola skor yang menyimpang pada item *polytomous* dalam bentuk skala EIS. Selain itu, penelitian ini juga menguji perbedaan nilai *person-fit* berdasarkan jenis kelamin dan etnik dari peserta tes. Seperti yang disebutkan sebelumnya bahwa studi terkait *person-fit* lebih berfokus pada item *dychotomous* dari pada *polytomous* sehingga studi *person-fit* untuk item *polytomous* masih perlu diuji mengingat alat ukur psikologi banyak yang menggunakan skala Likert. Selain itu, penelitian mengenai hubungan *person-fit* dengan jenis kelamin dan etnik masih belum dilakukan di Indonesia.

#### A. Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan permasalahan dalam penelitian ini adalah:

1. Bagaimanakah aplikasi *person fit* dengan statistik  $I_z$  dalam mendeteksi pola skor yang menyimpang pada item *polytomous*?
2. Apakah terdapat hubungan *person-fit* dengan jenis kelamin dan etnik?

#### B. Tujuan Penelitian

Tujuan penelitian ini adalah:



1. Untuk mengetahui kemampuan uji *person-fit* dengan statistik  $l_z$  dalam mendeteksi pola skor yang menyimpang pada skala model *polytomous*.
2. Untuk mengetahui hubungan *person-fit* dengan jenis kelamin dan etnik.

### C. Manfaat penelitian

1. Manfaat teoritis

Sebagai salah satu pengembangan dalam penggunaan *person-fit* dengan statistik  $l_z$  pada alat ukur *polytomous* dan hubungan *person-fit* dengan jenis kelamin dan etnik.

2. Manfaat praktis

Dapat digunakan sebagai acuan bagi penelitian selanjutnya terkait uji *person-fit* pada alat ukur *polytomous* dan hubungan *person-fit* dengan jenis kelamin dan etnik

### D. Penelitian terdahulu

1. Sebuah studi oleh Drasgow, F., Levine, M. V., & McLaughlin, M. E (1998) menggunakan tiga parameter logistik (3PL) model untuk membandingkan kinerja statistik *person fit* dan menyimpulkan bahwa statistik  $l_z$  termasuk yang paling mampu untuk mengidentifikasi perilaku menyimpang. Li & Olejnik (1997) membandingkan distribusi, statistik  $l_z$ ,  $ZU$ ,  $ZW$ ,  $2$ ,  $ECI z$  dan  $4 ECI z$  dalam kerangka model *Rasch* dan menemukan bahwa statistik  $l_z$  adalah yang paling kuat, karena mengidentifikasi dua pertiga dari *misfitting* pola skor-item.

2. Studi kedua oleh Meijer (1997) menggunakan simulasi data yang hasilnya menunjukkan bahwa ada penurunan substansial dalam validitas tes ketika tingkat *misfit* berat (yaitu, orang menebak jawaban yang benar semua item), persentase vektor respons *misfitting* adalah 15% atau lebih besar, dan uji validitas berada di tingkat moderat (yaitu, 0,3 hingga 0,4). Mengabaikan peserta tes yang terindikasi *misfit* dengan  $I_z$  menghasilkan peningkatan validitas.
3. Glas dan Dagohoy (2007) kemudian melakukan penelitian untuk menguji model IRT untuk *person-fit*. Model IRT yang diuji adalah *Generalized Partial Credit Model*, *Sequantial Model*, dan *Graded Response Model*. Studi ini dilakukan dengan studi simulasi dan dengan menggunakan sebuah *set data* dari *NEO Personality Inventory-Revised*. Penelitian ini menunjukkan hasil bahwa dari ketiga model yang diuji memberikan hasil yang sebanding dalam mengukur *person-fit*.
4. Schmitt, Chan, dan Sacco (1999) melakukan studi dengan menghubungkan *person-fit* dengan beberapa variabel, yaitu tes reaksi, motivasi mengerjakan tes, ketelitian, jenis kelamin, dan ras. Studi ini menunjukkan hasil bahwa peserta tes laki-laki memiliki nilai statistik *person-fit* yang negatif lebih tinggi dari perempuan. Peserta tes dengan ras Afrika-Amerika memiliki nilai statistik *person-fit* yang negatif lebih tinggi dari peserta tes kulit putih. Hal tersebut menunjukkan bahwa pola respon yang menyimpang lebih banyak ditunjukan oleh laki-laki dibanding perempuan, dan Afrika-Amerika dibanding kulit putih.

## BAB II

### KAJIAN TEORI

#### A. *Person-Fit*

##### 1. Pendekatan *Person-fit*

*Person-fit* merupakan salah satu model statistik pendeteksi pola respons yang atipikal atau *aberrant*. *Person-fit* adalah indeks statistik yang dapat digunakan sebagai ukuran langsung terhadap keakuratan pengukuran dan identifikasi pola-pola respon yang menyimpang berdasarkan model pengukuran yang dipakai. Skor tes yang didapatkan individu yang memiliki nilai *person-fit* yang rendah berpotensi untuk tidak akurat mengukur abilitas murni mereka.

Menurut Meijer dan Sijtsma (2001), metode *person-fit* merupakan suatu metode statistik yang dipergunakan untuk mengevaluasi ketidaksesuaian kinerja tes individu pada model IRT atau pola item-skor lainnya dalam sampel berupa sekelompok orang. Statistik *person-fit* dirancang untuk mengidentifikasi partisipan dengan pola respon item yang menyimpang (mengarah ke skor tes terlalu tinggi atau terlalu rendah), dan untuk membedakan partisipan tersebut dari partisipan dengan pola respon yang normal dan biasa (Karabatsos, 2003).

## 2. Statistik *Person-fit*

Terdapat beberapa statistik dalam *person-fit*. Di dalam Karabatsos (2003) statistik *person-fit* dijabarkan menjadi dua garis besar, yaitu *parametric* dan *non-parametric*, dengan pembagian statistik sebagai berikut:

- a. *Non-parametric*, merupakan statistik *person-fit* yang tidak didasari oleh model *Item Response Theory* (IRT). Terdapat 11 statistik yaitu  $G$ ,  $G^*$ ,  $r_{pbis}$ ,  $C$ ,  $MCL$ ,  $U3$ ,  $ZU3$ ,  $H^T$ ,  $A$ ,  $D$ ,  $E_i$ .
- b. *Parametric*, merupakan statistik *person-fit* yang berlandaskan model *Item Response Theory* (IRT). Terdapat 25 statistik, yaitu:  $U$ ,  $ZU$ ,  $\ln U$ ,  $W$ ,  $ZW$ ,  $\ln W$ ,  $ECI1$ ,  $ECI2$ ,  $ECI3$ ,  $ECI4$ ,  $ECI5$ ,  $ECI6$ ,  $ECI1_z$ ,  $ECI2_z$ ,  $ECI4_z$ ,  $ECI6_z$ ,  $l$ ,  $l_z$ ,  $M$ ,  $M_{(p-value)}$ , *Item-Grouping Person-fit Statistics*:  $D(\theta)$ ,  $l_{zm}$ ,  $UB$ ,  $ZUB$ , dan  $\ln UB$ .

Penelitian ini menggunakan salah satu statistic *person-fit* yaitu statistik  $l_z$ . Statistik  $l_z$  merupakan salah satu statistik *parametric* dalam *person-fit*. Statistik ini didasarkan pada indeks  $l_o$  menggunakan maksimum fungsi *likelihood*. Pola respons yang lebih konsisten dengan model IRT akan memiliki nilai  $l_o$  yang lebih tinggi, mencerminkan *person-fit* yang lebih baik. Nilai yang lebih rendah dari  $l_o$  mencerminkan *person-fit* yang buruk.

Statistik  $l_o$  mewakili tingkat kecocokan dari suatu pola respons dari sebuah  $\theta$ .  $l_o$  ini didefinisikan sebagai berikut (Levine & Rubin, 1979) dalam persamaan (1):

$$l_o(\theta_j) = \sum_{i=1}^n [u_{ij} \ln P_i(\theta_j) + (1 - u_{ij}) \ln Q_i(\theta_j)] \quad (1)$$



dimana  $n$  mengacu pada jumlah item dalam pengujian,  $u_{ij}$  mewakili respons dari peserta tes  $j$  terhadap item  $i$ ,  $P_i(\theta_j)$  adalah probabilitas dari respon terhadap item yang diberikan kepada peserta tes  $j$ . Persamaan ini mengasumsikan bahwa  $\theta$  seseorang tidak berubah selama tes diberikan kepada peserta tes dengan  $\theta$  tertentu.  $l_o$  menghasilkan satu nilai estimasi  $\theta$  pada setiap pola respon dari tiap peserta tes. Namun, Drasgow dkk. (1985) menetapkan bahwa distribusi  $l_o$  berubah sebagai fungsi dari  $\theta$ . Akibatnya, nilai  $l_o$  berubah tergantung pada lokasi peserta tes di kontinum  $\theta$ . Drasgow dkk, (1985) mengatasi ketergantungan kondisional ini dari  $l_o$  terhadap  $\theta$  dengan standarisasi  $l_o$ . Standarisasi tersebut dirumuskan dalam persamaan (2), (3) dan (4) berikut:

$$l_z = \frac{l_o - E(l_o)}{[\text{Var}(l_o)]^{1/2}} \quad (2)$$

dimana,

$$E(l_o) = \sum_{i=1}^n \{P_i(\theta_j) \ln P_i(\theta_j) + [1 - P_i(\theta_j)] \ln [1 - P_i(\theta_j)]\}, \quad (3)$$

$$\text{Var}(l_o) = \sum_{i=1}^n P_i(\theta_j) [1 - P_i(\theta_j)] \left\{ \ln \left[ \frac{P_i(\theta_j)}{1 - P_i(\theta_j)} \right] \right\}^2 \quad (4)$$

Hasil tes yang ditunjukkan oleh nilai negatif yang besar dari statistik  $l_z$  ( $< -2.00$ ) maka dapat diidentifikasi sebagai pola respon *aberrant* (Schmitt, Chan, & Sacco., 1999)

### 3. Pola Respon *Aberrant*

Pola respon yang menyimpang (*aberrant*) merupakan pola respon peserta tes yang dianggap atipikal (mengarah ke skor terlalu tinggi atau terlalu rendah). Respon *aberrant* ini disebabkan oleh perilaku yang menyimpang pada peserta tes saat mengerjakan tes. Respon *aberrant* dapat mempengaruhi akurasi hasil tes berdasarkan kemampuan peserta tes. Berikut jenis peserta tes yang dapat memberikan pola respon *aberrant* (Meijer, 1996).

- a. Respon *aberrant* yang secara artifisial meningkatkan estimasi kemampuan peserta tes, dibagi menjadi dua yaitu *cheaters* dan *lucky guessers*. *Cheaters* merujuk pada peserta tes yang melakukan kecurangan sehingga memberikan respon yang benar pada item-item yang sulit. *Lucky guessers* merujuk peserta tes yang secara tidak sengaja menebak respon beberapa item sulit dengan benar
- b. Respon *aberrant* yang secara artifisial menurunkan estimasi kemampuan peserta tes, dibagi menjadi tiga yaitu *flumbers*, *plodders*, dan *language deficiency*. *Fumblers* merujuk pada peserta tes yang bingung di awal pengerjaan tes yang kemudian memungkinkan peserta tes memberikan respon yang salah pada item-item pertama. *Plodders* merujuk pada peserta tes yang tidak menyelesaikan tes hingga tuntas sehingga memberikan respon salah pada item-item terakhir. *Language deficiency* merujuk pada peserta tes yang tidak

memahami penggunaan bahasa dalam tes sehingga memberikan respon yang salah pada item-item yang relatif mudah.

c. Respon *aberrant* yang secara artifisial dapat meningkatkan ataupun menurunkan estimasi kemampuan peserta tes, terbagi menjadi tiga, yaitu *random guessers*, *creatives*, dan *palm beachers*. *Random guessers* merujuk pada peserta tes yang merespon tanpa memikirkan konten dari item (jawaban yang benar dan salah yang tidak diharapkan). *Creatives* merujuk pada peserta tes yang mengartikan item dengan cara yang kreatif (jawaban benar dan salah yang tidak diharapkan). *Palm beachers* merujuk pada peserta tes yang salah mengisi lembar jawaban mereka.

#### **B. Model *Polytomous* dalam IRT**

Pendekatan *Item Response Theory* (IRT) merupakan pendekatan alternatif yang dapat digunakan dalam menganalisis suatu tes. Berdasarkan kategori respon, model pengukuran IRT dibedakan menjadi model IRT *dychotomous* dan model IRT *polytomous*. Model IRT dikotomi digunakan pada tes yang item-nya memiliki dua kategori respon. Model *polytomous* adalah untuk item yang memiliki lebih dari dua kemungkinan skor. Contoh yang paling umum adalah item jenis Likert.

Salah satu model dalam *Item Response Theory* (IRT) untuk data politomus adalah *Graded Response Model* (GRM), yang merupakan sistem penskoran dimana tingkat kesukaran tiap kategori pada butir tes disusun secara

berurutan sehingga jawaban peserta tes haruslah terurut dari kategori yang rendah hingga kategori yang tinggi. GRM sangat tepat digunakan untuk butir yang memiliki respons kategorikal seperti skala likert. GRM merupakan perluasan dari model 2-PL. Model 2-PL dapat digambarkan dengan persamaan (5), dan (6):

$$P_i(\theta) = \left( \frac{1}{1 + e^{-a(\theta - b)}} \right) = \left( \frac{1}{1 + e^{-L}} \right) \quad (5)$$

$$Q = 1 - P_i(\theta) \quad (6)$$

Dimana  $P$  merupakan probabilitas menjawab benar setiap peserta tes,  $a$  merupakan parameter *item discrimination*,  $b$  merupakan parameter *item difficulty*, dan  $e$  merupakan nilai konstan dengan nilai 2.718.

Dikarenakan GRM dikembangkan berdasarkan model 2-PL, maka GRM juga berkaitan dengan dua parameter item, yaitu:

- a. Parameter lokasi ( $\beta_j$ )

Parameter ini dikenal dengan *item difficulty* atau tingkat kesukaran item. Parameter ini berkaitan dengan kemampuan individu atau posisi *latent trait* dan probabilitas individu memilih suatu kategori jawaban pada item.

- b. Parameter *slope* ( $\alpha_j$ )

Parameter  $\alpha_j$  merupakan *item discrimination* atau daya pembeda item. Tujuan dari kebanyakan pengukuran psikologis adalah memberikan informasi mengenai perbedaan individu (*individual differences*) dalam hal konstruk yang diukur tes atau dari kriteria eksternal yang ingin diprediksi oleh

alat tes tersebut (Crocker & Algina, 1986). Oleh karena itu sebuah tes perlu memiliki daya diskriminasi. Pengertian dari daya diskriminasi adalah derajat dimana item dapat membedakan responden dengan skor tinggi dan skor rendah (Cohen & Swerdlik, 2005). Semakin tinggi nilai  $\alpha_j$  maka semakin baik item tersebut. Selanjutnya, fungsi GRM dirumuskan dalam persamaan (7), dan (8):

$$P_{ic}(\theta) = P_{ic}^*(\theta) - P_{i(c+1)}^*(\theta), \quad (7)$$

dimana,

$$P_{ic}^*(\theta) = \left( \frac{1}{1 + e^{-\alpha_i(\theta - \beta_{ic})}} \right) \quad (8)$$

Dimana,  $P_{ic}$  adalah probabilitas peserta tes memilih pilihan respon  $c$ , dan  $P_{ic}^*$  adalah probabilitas peserta tes memilih pilihan respon setelah  $c$ ,  $\alpha_i$  merupakan parameter *item discrimination* pada item  $i$ ,  $\beta_{ic}$  merupakan parameter *item difficulty* pada item  $i$  untuk pilihan respon  $c$ , dan  $e$  merupakan nilai konstan dengan nilai 2.718.

### C. *Person-Fit* dengan Jenis Kelamin dan Etnik

Menurut Hungu (2007), jenis kelamin (*sex*) adalah perbedaan antara perempuan dengan laki-laki secara biologis sejak seseorang lahir. Lebih lanjut, jenis kelamin memungkinkan perbedaan pada laki-laki dan perempuan selain perbedaan biologis, misalnya perbedaan peran dan status sosial.

Pengertian etnis atau suku adalah kategori sosial yang digunakan masyarakat untuk membedakan suatu kelompok tertentu yang mempunyai ciri-



ciri berbeda dengan kelompok lain (Phinney, 1992). Identifikasi sebuah kelompok etnis menurut Phinney (1992) diawali oleh perasaan sadar akan kesamaan budaya.

Sebuah studi (Schmitt, Chan, & Sacco., 1999) mengembangkan pengukuran *person-fit* yang dihubungkan dengan beberapa variabel lainnya. Diantaranya adalah reaksi tes, motivasi mengerjakan tes, ketelitian, jenis kelamin, dan ras. Penelitian ini dilakukan dengan subjek dari dua kelompok ras, yaitu ras kulit putih dan Afrika-Amerika. Nilai  $I_z$  yang mengindikasikan adanya pola respon yang menyimpang diasumsikan sebagai manifestasi dari motivasi yang rendah dalam mengerjakan tes. Dijelaskan pula bahwa  $I_z$  diasumsikan memiliki hubungan dengan reaksi tes, motivasi mengerjakan tes, dan ketelitian.

Asumsi selanjutnya dalam penelitian Schmitt, Chan, dan Sacco (1999) menyatakan bahwa peserta tes dengan jenis kelamin dan ras tertentu memiliki reaksi tes negatif, motivasi dan ketelitian rendah dalam mengerjakan tes. Berdasarkan hasil penelitiannya, ditemukan bahwa peserta tes laki-laki memiliki nilai  $I_z$  negatif yang lebih tinggi dari pada peserta tes perempuan. Dan peserta tes Afrika-Amerika memiliki nilai  $I_z$  negatif yang lebih tinggi dari pada peserta tes kulit putih. Hal tersebut menunjukkan bahwa pola respon yang menyimpang lebih banyak ditunjukkan oleh laki-laki dibanding perempuan, dan Afrika-Amerika dibanding kulit putih. Penelitian serupa masih belum dilakukan di Indonesia, sehingga penelitian ini lebih lanjut mengasumsikan bahwa pola respon yang menyimpang dipengaruhi oleh jenis kelamin dan etnisitas peserta tes.

Horner (1975) menyimpulkan bahwa kebanyakan perempuan memiliki tingkat motivasi yang lebih tinggi daripada laki-laki dalam hal akademik. Berdasarkan penelitian lain, diketahui bahwa kecurangan akademik lebih sering terjadi pada anak laki-laki daripada anak perempuan. Menurut banyak penulis (Honny, Gadbury-Amyot, Overman, Wilkins, Petersen, 2010; Kobayashi; Fukushima, 2012; Salleh, Hamid, Alias, Ismail,; Yusoff, 2011; Saulsbury, Brown, Heyliger, & Beale 2011), anak perempuan cenderung tidak melakukan kecurangan akademik di sekolah karena mereka biasanya membangun rasa tanggung jawab yang lebih kuat dibanding anak laki-laki. Kedua hal ini dapat memicu peola respon yang menyimpang saat pengerjaan tes.

Uji beda *person-fit* berdasarkan jenis kelamin dan etnik dapat dilakukan dengan uji *multivariate anova*, uji ini digunakan untuk menguji banyak kelompok sampel yang melibatkan klasifikasi ganda (lebih dari satu variabel dependen).

#### **D. Identitas Etnik**

##### **1. Definisi Identitas Etnik**

Phinney (1992) menyatakan bahwa identitas etnis merupakan konstruksi yang kompleks yang mencakup komitmen dan perasaan lekat pada suatu kelompok, evaluasi positif tentang kelompoknya, adanya ketertarikan dan pengetahuan tentang kelompok, serta keterlibatan dalam aktivitas sosial dari kelompok. Phinney juga menjelaskan identitas etnis sebagai suatu identitas seseorang atau *sense of self* sebagai seorang anggota dari sebuah kelompok etnis

dan pemikiran, persepsi serta perasaan yang dirasakan seseorang sebagai bagian dari anggota kelompok tersebut.

Identitas etnik merupakan sesuatu yang dinamis, yang berarti bahwa identitas etnis dapat berubah sepanjang waktu dan juga konteks, dan harus disesuaikan dengan variasi dan pembentukannya (Phinney, 1992). Identitas etnis merupakan suatu konsepsi diri terbentuk sebagai hasil dari pembentukan. Perkembangan identitas etnis merupakan pergerakan individu untuk mengidentifikasi nilai-nilai budaya, perilaku, kepercayaan, dan tradisinya (Chavez & Guido-Dibrito, 1999).

## 2. Dimensi Identitas Etnik

Umaña-Taylor, Yazedjian, & Bámaca-Gómez (2004) melengkapi teori Phinney & Ong (2007) dengan memasukkan dimensi afirmasi (positif dan negatif) setelah terdapat dua dimensi sebelumnya yaitu eksplorasi dan komitmen.

### a. Eksplorasi

Eksplorasi merupakan proses pencarian informasi dan pengalaman yang relevan dengan etnik individu. Proses ini terjadi berkelanjutan sepanjang hidup individu.

### b. Komitmen

Komitmen merupakan perasaan terikat dalam diri individu terhadap etniknya. Komitmen yang ada pada diri individu terhadap etniknya kemudian menjadikan individu merasa jadi bagian dan mengambil peran dalam kegiatan etniknya.

c. Afirmasi

Afirmasi merujuk pada penilaian atau evaluasi individu terhadap etnik asalnya.

3. Skala Identitas Etnik

Alat ukur identitas etnik yang digunakan adalah skala EIS oleh Umaña-Taylor (2004) yang telah ditransadaptasi oleh Paramita (2014) dan diuji ulang oleh Ningsih (2017). Skala ini terdiri dari 17 item yang mengukur tiga dimensi yaitu eksplorasi, komitmen, dan afirmasi. Skala ini menggunakan model skala Likert. *Blueprint* skala dapat dilihat pada lampiran.

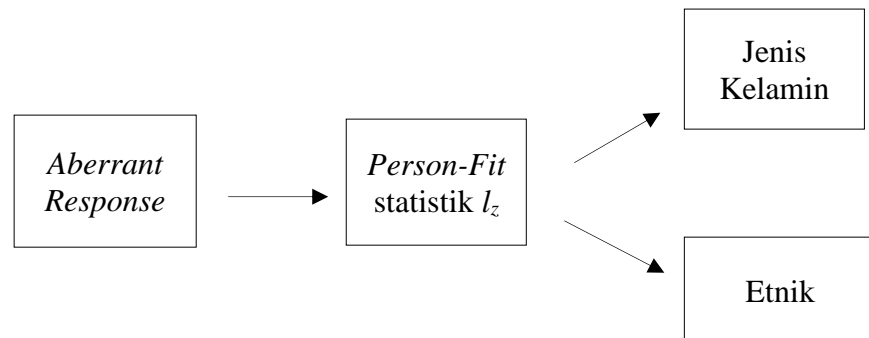
## E. Variabel Penelitian

Adapun variabel penelitian ini dapat dilihat pada Tabel 1:

Tabel 1.  
*Variabel penelitian*

Statistik $l_z$	Jenis kelamin	Etnik
Salah satu statistik yang banyak digunakan dalam pengaplikasian <i>person-fit</i> adalah statistik $l_z$ yang merupakan pengembangan dari $l_o$ . Statistik $l_z$ akan digunakan dalam penelitian ini untuk mendeteksi peserta tes yang menunjukkan pola respon <i>aberrant</i> .	Jenis kelamin adalah perbedaan antara perempuan dengan laki-laki secara biologis.	Etnik adalah suatu golongan manusia yang terikat oleh kesadaran dan identitas akan kesatuan kebudayaan dan ciri khusus. Penelitian ini mengasumsikan terdapat perbedaan nilai statistik $l_z$ pada masing-masing etnik peserta tes.

## F. Kerangka Pemikiran



Gambar 1. Kerangka Penelitian

Kecenderungan manipulasi dalam mengisi skala dapat menimbulkan pola respon yang menyimpang dan akan mempengaruhi akurasi sebuah asesmen. Guna mengidentifikasi pola respon menyimpang tersebut, dapat diatasi dengan uji *person-fit*. Salah satu statistik *person-fit* yang disarankan adalah  $l_z$ . Uji *person-fit* dapat dilakukan setelah melakukan analisis terhadap probabilitas peserta tes, posisi *latent trait* peserta tes, dan item pada tes yang dapat diukur dengan GRM untuk alat ukur *polytomous*. Setelah itu dapat dilakukan uji beda nilai *person-fit* berdasarkan jenis kelamin dan etnik untuk melihat perbedaan nilai statistik  $l_z$  pada peserta tes laki-laki dan perempuan, serta berdasarkan etnik masing-masing peserta tes.

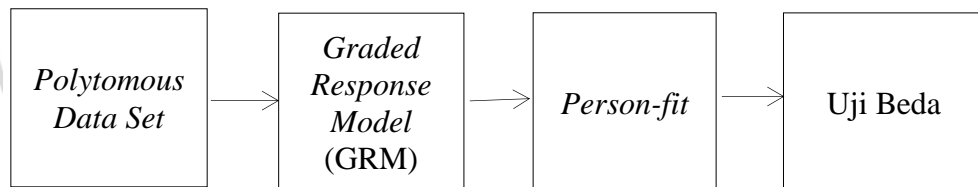


### BAB III

#### METODE PENELITIAN

##### A. Desain Penelitian

Metodologi yang digunakan pada penelitian ini adalah kuantitatif psikometrik dengan melakukan uji *person-fit* dengan statistik  $l_z$  pada hasil tes dari EIS yang telah dianalisis menggunakan pendekatan IRT dengan model *Graded Response Model* (GRM). Selanjutnya akan dilakukan uji beda *person-fit* berdasarkan jenis kelamin dan etnik untuk melihat perbedaan nilai statistik  $l_z$  pada peserta tes laki-laki dan perempuan, serta berdasarkan etnik masing-masing peserta tes. Bagan desain penelitian dapat dilihat pada Gambar 1.



Gambar 2. Desain penelitian *person-fit* pada item *polytomous*

##### B. Fokus Penelitian

Penelitian ini difokuskan untuk menguji keberfungsian metode *person-fit* dengan statistik  $l_z$  untuk mendeteksi pola respon yang menyimpang dengan model item *polytomous*. Lebih lanjut, penelitian ini juga difokuskan untuk menguji perbedaan nilai *person-fit* berdasarkan jenis kelamin dan etnik.

### C. Definisi Operasional

*Person-fit* merupakan suatu metode statistik yang dipergunakan untuk mendeteksi adanya pola respon yang menyimpang (*aberrant*) dalam suatu *set data*. Penelitian ini menggunakan statistik  $I_z$  karena lebih mudah untuk diinterpretasi dan terbukti memiliki kinerja yang lebih baik berdasarkan beberapa studi dibandingkan dengan statistik lainnya dalam mendeteksi pola respon menyimpang.

Jenis kelamin (*sex*) adalah perbedaan antara perempuan dengan laki-laki secara biologi. Etnik adalah suatu golongan manusia yang terikat oleh kesadaran dan identitas akan kesatuan kebudayaan dan ciri khusus. Penelitian ini akan menguji perbedaan nilai *person-fit* berdasarkan jenis kelamin dan etnik untuk mengetahui perbedaan nilai statistik  $I_z$  pada peserta tes laki-laki dan perempuan, serta etnik masing-masing peserta tes.

### D. Subjek Penelitian

Populasi penelitian ini adalah remaja akhir perempuan dan laki-laki berusia 15 sampai dengan 18 tahun di kota Malang. Sedangkan sampel penelitian ini adalah remaja akhir yang berusia antara 15 sampai dengan 18 tahun di SMA Katolik Kolese Santo Yusup Malang yang berjumlah 560 siswa. Karakteristik subjek mengacu pada teori identitas etnik dimana menurut Phinney (1992), identitas etnik merupakan aspek dari identitas yang perlu diperhatikan sebagai sesuatu yang penting selama masa remaja, dan identitas etnik mempengaruhi perkembangan psikologis individu. Teknik pengambilan sampel yang digunakan adalah *non probability sampling*,

dimana sampel tidak memiliki peluang yang sama untuk menjadi subjek penelitian karena sampel ditentukan oleh pihak sekolah.

#### E. Tahap pelaksanaan penelitian

Pelaksanaan penelitian ini terdiri dari 3 tahap, yaitu tahap persiapan, tahap pengumpulan data dan tahap analisis data. Tahapan tersebut kemudian dijabarkan sebagai berikut:

- a. Peneliti menyiapkan skala EIS (Umaña-Taylor, dkk., 2004) yang telah ditransadaptasi oleh Paramita (2014).
- b. Peneliti membuat surat izin penelitian
- c. Peneliti meminta izin kepada pihak terkait untuk melakukan penelitian..
- d. Peneliti melakukan penggalan data kepada 560 subjek yang berusia 15 – 18 tahun di SMA Katolik Kolese Santo Yusup Malang.
- e. Peneliti memastikan data yang telah dikumpulkan layak untuk digunakan.
- f. Peneliti memberikan skor pada setiap item sesuai dengan kategori jawaban yang dipilih peserta tes serta membuat tabulasi data
- g. Peneliti menganalisis data yang diperoleh menggunakan *software R-Studio* versi 3.3.2 dengan menggunakan metode IRT – *Graded Response Model* (GRM), kemudian menganalisis hasil dari GRM.
- h. Peneliti menganalisis data dari GRM dengan *person-fit* menggunakan statistik  $l_z$  dengan *software* yang sama menggunakan *package PerFit*.
- i. Peneliti menganalisis uji beda *person-fit* berdasarkan jenis kelamin dan etnik.
- j. Peneliti melakukan interpretasi data.

### A. Teknik Analisis Data

Penelitian ini menggunakan beberapa teknik perhitungan yang termasuk ke dalam model analisis IRT. Aplikasi *software* yang digunakan untuk menganalisis data adalah *R* versi 3.3.2.

#### 1. *Graded Response Model* (GRM)

GRM merupakan model perhitungan yang mengadaptasi model 2-PL yang diterapkan dalam item *polytomous*. sistem penskoran dimana tingkat kesukaran tiap kategori pada butir tes disusun secara berurutan sehingga jawaban peserta tes haruslah terurut dari kategori yang rendah hingga kategori yang tinggi. Cara perhitungan GRM dapat dilihat pada persamaan (6), dan (7).

#### 2. *Person-fit* dengan statistik $l_z$

*Person-fit* ditujukan untuk mengidentifikasi pola respon menyimpang dari peserta tes dari hasil analisis data item *polytomous* yang telah dilakukan sebelumnya. *Person-fit* pada penelitian ini menggunakan statistik  $l_z$  yang dapat dilihat pada persamaan (2), (3), dan (4).

#### 3. Uji beda

Uji beda ditujukan untuk mengetahui perbedaan jenis kelamin dan etnik dari peserta tes terhadap nilai *person-fit*. Uji beda ini dilakukan dengan *multivariate anova*.

## BAB IV

### HASIL DAN PEMBAHASAN

#### A. Gambaran Umum Subjek Penelitian

Subjek penelitian ini merupakan remaja dengan rentang usia 15 hingga 18 tahun yang sedang bersekolah di SMA Katolik Kolese Santo Yusup Malang. Jumlah subjek yang digunakan pada penelitian ini sebanyak 560 siswa. Subjek penelitian yang digunakan terdiri dari siswa kelas 10 dan 11. Kategorisasi subjek pada penelitian ini didasarkan pada jenis kelamin, usia, kelas, dan etnik. Kategorisasi subjek penelitian dapat dilihat pada Tabel 2.

Tabel 2.  
*Kategorisasi subjek penelitian*

Kategorisasi Subjek Penelitian	Kategori	Jumlah	Persentase (%)
Jenis Kelamin	Laki-laki	232	41.43
	Perempuan	328	55.57
Usia	15	313	55.89
	16	200	35.72
	17	39	6.96
	18	8	1.43
Kelas	X	320	57.14
	XI	240	42.86
Etnik	Bali	2	.35
	Batak	1	.18
	Betawi	1	.18

(lanjut)



Tabel 2.  
*Kategorisasi subjek penelitian*

Kategorisasi Subjek Penelitian	Kategori	Jumlah	(lanjutan)
			Persentase (%)
Etnik	Bima	1	.18
	Bugis	1	.18
	Dayak	4	.71
	Jawa	372	66.43
	Madura	2	.35
	Melayu	2	.35
	Minang	1	.18
	Tionghoa	169	30.20
	Toraja	1	.18
	Sasak	1	.18
	Sunda	2	.35

## B. Hasil Penelitian

Penelitian ini bertujuan untuk mengetahui kemampuan uji *person-fit* dengan statistik  $I_z$  dalam mendeteksi pola respon yang menyimpang (*aberrant*) pada skala *polytomous*. Lebih lanjut, penelitian ini bertujuan pula untuk mengetahui perbedaan nilai *person-fit* berdasarkan jenis kelamin dan etnik. Adapun skala *polytomous* yang digunakan dalam penelitian ini merupakan *Ethnic Identity Scale* (EIS) yang telah ditransadaptasi oleh Paramita (2014) dan telah diuji kembali oleh Ningsih (2017). EIS terdiri dari 17 item yang mencakup tiga dimensi, yaitu eksplorasi, komitmen, dan afirmasi. Berikut adalah penjabaran dari hasil analisis data yang diperoleh.

### 1. Parameter Item dengan *Graded Response Model*

Penelitian ini menggunakan salah satu model IRT dalam mengestimasi parameter item, yaitu *Graded Response Model* (GRM). GRM digunakan

untuk mengukur parameter  $\alpha_j$  dan  $\beta_j$ . Nilai parameter item dari data set EIS dapat dilihat pada Tabel 3.

Tabel 3.

*Parameter item dengan Graded Response Model*

Item	$\beta_1$	$\beta_2$	$\beta_3$	$\alpha$
1	-2.434	-1.466	-.382	2.236
2	-2.679	-1.591	-.458	2.004
3	-2.440	-1.387	-.389	1.951
4	-2.494	-1.303	-.390	2.168
5	-2.679	-1.657	-.367	2.236
6	-2.562	-1.678	-.462	1.844
7	-2.673	-1.513	-.460	2.141
8	-2.743	-1.516	-.405	2.078
9	-2.829	-1.657	-.425	1.875
10	-2.269	-1.308	-.389	1.941
11	-2.486	-1.421	-.335	1.974
12	-2.560	-1.624	-.469	2.202
13	-2.564	-1.499	-.398	1.667
14	-2.560	-1.437	-.321	2.427
15	-2.748	-1.515	-.452	2.358
16	-2.525	-1.490	-.436	2.339
17	-2.438	-1.604	-.304	2.043

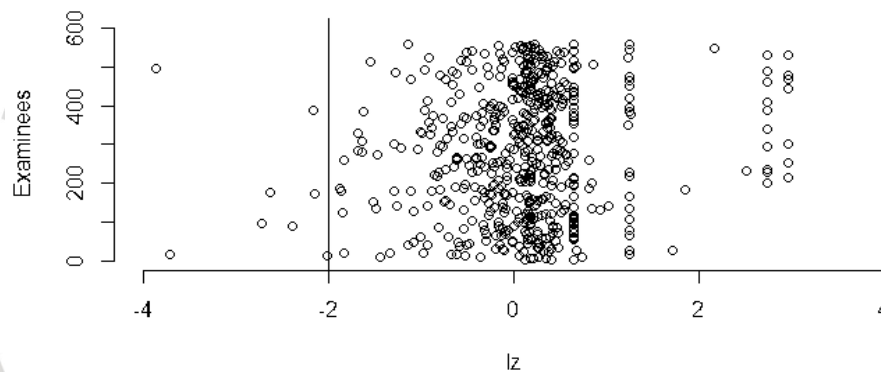
Nilai *extremity parameter* adalah nilai *latent trait* ( $\theta$ ) yang memiliki probabilitas 50:50 untuk memilih suatu kategori jawaban. Berdasarkan EIS yang memiliki 4 kategori jawaban, maka terdapat 3 nilai *extremity parameter* untuk tiap-tiap item.

Menurut Baker (2001), nilai  $\beta_j$  berada pada  $-3 \leq \beta_j \leq 3$  dan  $a_j$  berada pada  $0 \leq a_j \leq \infty$ , parameter  $a_j$  yang berada pada rentang .65 hingga 1.70 sudah dapat termasuk ke dalam parameter item diskriminasi yang baik. Menurut hasil analisis GRM yang dilakukan, dapat diketahui bahwa keseluruhan item pada skala EIS memiliki estimasi parameter  $a_j$  dengan nilai  $0 \leq a_j \leq \infty$ , sehingga seluruh item pada EIS memiliki parameter *item*

*discrimination* yang baik. Hal tersebut menunjukkan bahwa item-item pada EIS mampu membedakan individu dengan *latent trait* tinggi dan rendah.

## 2. Skor *Person-Fit* Statistik $l_z$

Penelitian ini menggunakan salah satu statistik *parametric* dari *person-fit*, yaitu statistik  $l_z$  untuk mendeteksi pola respon *aberrant* pada *data set*.  $l_z$  dianggap unggul dalam mendeteksi *aberrant* dibanding kebanyakan statistik *person-fit* yang lain. Hasil dari uji statistik  $l_z$  dapat dilihat pada Gambar 3.



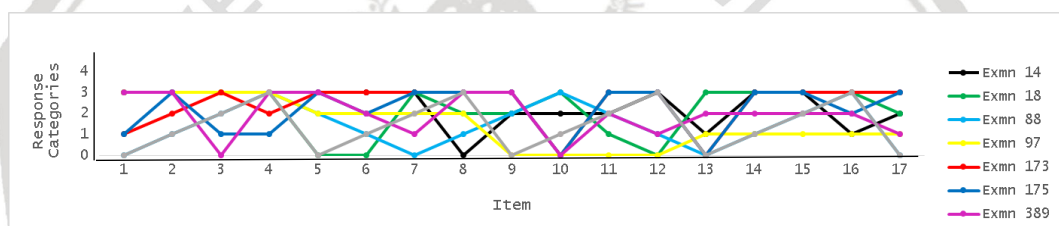
Gambar 3.  $l_z$  value peserta tes

Subjek yang memiliki  $l_z$  value ( $<-2.00$ ) dianggap sebagai *aberrance*. Diketahui dari hasil uji statistik  $l_z$  bahwa dari 560 subjek terdeteksi sebanyak 8 (1.42%) subjek yang memiliki  $l_z$  value ( $>-2.00$ ) dan dianggap sebagai *aberrance*. Sedangkan 552 sisanya sebagai *non- aberrance* karena memiliki  $l_z$  value ( $>-2.00$ ).  $l_z$  value subjek yang dianggap *aberrance* dapat dilihat pada Tabel 3.

Tabel 4.  
 $l_z$  value yang terdeteksi sebagai aberrance

Examinees	$l_z$ value
14	-2.0207
18	-3.714
88	-2.3973
97	-2.7324
173	-2.1501
175	-2.6325
389	-2.1725
496	-3.8771

Adapun pola respon atau skor yang diberikan peserta tes yang terdeteksi *aberrance* dapat dilihat pada Gambar 4.



Gambar 4. Pola respon peserta tes *aberrance*

Berdasarkan pola respon yang diberikan oleh peserta tes yang terdeteksi sebagai *aberrance*, diketahui bahwa dari semua peserta tes yang dianggap sebagai *aberrance* menjawab item-item pada EIS dengan tidak konsisten. Kedelapan peserta tes tersebut memilih pilihan jawaban dengan skor tertinggi yang mengindikasikan identitas etnik yang tinggi pada beberapa item. Namun, pada beberapa item lainnya peserta tes yang dianggap *aberrance* ini juga memilih kategori terendah yang mengindikasikan identitas etnik yang rendah.

Status tinggi rendahnya identitas etnik peserta tes ini berkaitan dengan *latent trait* ( $\theta$ ) yang dimiliki tiap peserta tes. *Latent trait* ( $\theta$ ) peserta tes yang terdeteksi sebagai *aberrance* dapat dilihat pada Tabel 4.

Tabel 5.  
*Latent trait* ( $\theta$ ) peserta tes yang terdeteksi sebagai *aberrance*

<i>Examinees</i>	$\theta$
14	-.180
18	-.378
88	-1.421
97	-1.392
173	.0001
175	-.427
389	-.743
496	-1.389

Peserta tes yang terdeteksi sebagai *aberrance* dianggap tidak konsisten dengan estimasi *latent trait* ( $\theta$ ) mereka (Schmitt, Chan, & Sacco., 1999). *Aberrance* memberikan respon yang menyimpang yang tidak sepenuhnya merepresentasikan *latent trait* ( $\theta$ ) mereka yang sesungguhnya.

### 3. Uji beda *Person-Fit* Berdasarkan Jenis kelamin dan Etnik

Lebih lanjut, penelitian ini menganalisis perbedaan  $l_z$  value dari masing-masing subjek berdasarkan jenis kelamin dan etnik mereka. Hasil uji ANOVA untuk uji beda  $l_z$  value dengan jenis kelamin dan etnik subjek dapat dilihat pada Tabel 6.

Tabel 6.  
*Uji beda  $l_z$  value berdasarkan jenis kelamin dan etnik peserta tes.*

	<i>F value</i>	<i>Pr(&gt;F)</i>
<i>Sex</i>	3.6649	.056
<i>Ethnic</i>	9.0324	.002
<i>Ethnic:Sex</i>	0.8915	.345



Berdasarkan uji ANOVA diketahui bahwa  $p = .056$  untuk perbedaan berdasarkan jenis kelamin. Nilai  $p$  (.056) menunjukkan perbedaan yang tidak signifikan antara jenis kelamin peserta tes terhadap  $I_z$  value. Sehingga dapat diketahui bahwa tidak terdapat perbedaan antara peserta tes laki-laki dan perempuan dalam kecenderungan menunjukkan respon yang menyimpang saat pengerjaan tes.

Selanjutnya, dapat diketahui nilai  $p = .002$  untuk perbedaan  $I_z$  value berdasarkan etnik peserta tes. Nilai  $p$  (.002) menunjukkan terdapatnya perbedaan yang signifikan antar etnik. Hal ini menunjukkan bahwa terdapat etnik tertentu yang lebih memungkinkan dalam menunjukkan pola respon *aberrant* dibanding etnik lainnya. Adapun nilai *mean* dari  $I_z$  value tiap jenis kelamin dan etniknya dapat dilihat pada Tabel 7.

Tabel 7.

*Perbandingan mean  $I_z$  value berdasarkan jenis kelamin dan etnik.*

<i>Ethnic</i>	<i>Male</i>	<i>Female</i>
Jawa	.308	.284
Tionghoa	-.069	-.019

Penelitian ini menggunakan subjek sebagai peserta tes dengan 14 jenis etnik yang berbeda. Berdasarkan 14 etnik peserta tes, diketahui bahwa etnik dengan etnik yang sama dengan jumlah yang mendominasi adalah etnik Jawa dan Tionghoa. Apabila membandingkan nilai *mean* dari  $I_z$  value antara kedua etnik tersebut, maka dapat diketahui bahwa peserta tes dari etnik Tionghoa memiliki nilai *mean* yang terendah.

### C. Pembahasan

Penelitian ini dilakukan untuk mengetahui kemampuan statistik  $I_z$  dalam mendeteksi *aberrant*, serta untuk mengetahui perbedaan *person-fit* berdasarkan jenis kelamin dan etnik dari subjek penelitian. Subjek penelitian adalah remaja akhir dengan rentang usia 15 sampai 18 tahun yang menempuh pendidikan di SMA Katolik Kolese Santo Yusup Malang. Subjek penelitian dengan jenis kelamin perempuan berjumlah 328 orang, dan laki-laki berjumlah 232.

Subjek penelitian ini berasal dari etnik yang beragam, namun etnik dengan jumlah yang cukup besar adalah etnik Jawa dan Tionghoa. Alat ukur yang digunakan pada penelitian ini adalah *Ethnic Identity Scale* (EIS) oleh Umaña-Taylor, Yazedjian, dan Bámaca-Gómez (2004). EIS yang digunakan pada penelitian ini telah ditransadaptasi oleh Paramita (2014), dan diuji ulang menggunakan *delayed alternate form* oleh Ningsih (2017).

EIS sendiri merupakan salah satu skala yang termasuk ke dalam golongan *low stake test*, yaitu tes yang tidak memiliki konsekuensi besar bagi peserta tes. Skala yang termasuk *low stake test* cenderung akan dikerjakan dengan tidak serius oleh peserta tes. *Low stake test* telah diasumsikan dan dalam beberapa kasus terbukti terkait dengan penurunan dalam motivasi dan kinerja saat mengerjakan tes (Wise & DeMars, 2005). Rendahnya motivasi dan kinerja peserta tes dapat memicu peserta tes berperilaku menyimpang saat mengerjakan tes. Hal ini kemudian dapat membuat peserta tes memiliki pola respon *aberrant*.

Berdasarkan hasil analisis GRM yang telah dilakukan, dapat diketahui nilai *extremity parameter* antar kategori jawaban pada EIS, serta nilai *item discriminant* ( $a$ ) pada masing-masing item. Terdapat tiga kelompok *extremity parameter* yang didapatkan (EIS memiliki 4 kategori jawaban). Nilai *extremity parameter* menunjukkan subjek dengan *latent trait* tertentu yang memiliki probabilitas 50:50 untuk memilih kategori tertentu.

Menurut Baker (2001), nilai  $\beta_j$  berada pada  $-3 \leq \beta_j \leq 3$  dan  $a_j$  berada pada  $0 \leq a_j \leq \infty$ , parameter  $a_j$  yang berada pada rentang .65 hingga 1.70 sudah dapat termasuk ke dalam parameter item diskriminasi yang baik. Parameter  $a_j$  pada *data set* memiliki nilai ( $>1.70$ ) pada keseluruhan item. Artinya, seluruh item pada EIS mampu membedakan individu dengan *latent trait* tinggi dan rendah dan merupakan item yang bagus.

Selanjutnya, dilakukan uji *person-fit* pada *data set* yang didapat. Uji *person-fit* merupakan metode statistik yang dilakukan untuk mengevaluasi pola respon peserta tes terhadap suatu tes (Meijer & Sijsma, 2001). Berdasarkan hasil uji statistik *person-fit* dengan statistik  $l_z$ , didapati hasil bahwa terdapat 8 (1.42%) subjek dari 560 subjek yang terdeteksi sebagai *aberrance*. Subjek yang terdeteksi sebagai *aberrant* ini dilihat dari nilai  $l_z$  value yang ( $-2.00$ ), sedangkan *non aberrance* memiliki  $l_z$  value ( $>-2.00$ ).

Apabila membandingkan jumlah *aberrance* dan *non-aberrance*, maka diketahui bahwa jumlah *aberrance* yang terdeteksi terbilang kecil. Meskipun  $l_z$  teruji sebagai statistik *person-fit* yang akurat, perlu diketahui pula bahwa  $l_z$  memiliki keterbatasam dalam mendeteksi pola respon

aberrant. Secara umum, akurasi statistik  $l_z$  berkaitan dengan panjang tes, semakin pendek sebuah tes atau semakin sedikit item pada tes maka akurasi  $l_z$  semakin menurun (Schmitt, Chan, & Sacco., 1999). Terdapat penelitian terdahulu mengenai *person-fit* yang dilakukan oleh Schmitt, Chan, dan Sacco (1999). Subjek penelitian Schmitt, Chan, dan Sacco (1999) mengerjakan beberapa jenis tes, yaitu tes kognitif yang terbagi menjadi tiga macam, *vocabulary* yang terdiri dari 38 item, *grammar* 22 item, and STP 36 item. Selanjutnya, subjek penelitian juga mengerjakan tes *NEO-Five Factor Inventory* sebanyak 60 item. Mereka juga diberikan tes reaksi yang terdiri dari 12 item, serta tes motivasi yang terdiri dari 3 item. Secara keseluruhan, subjek penelitian Schmitt, Chan, dan Sacco (1999) diberikan sebanyak 171 item.

Penelitian ini menggunakan EIS untuk tes, dimana EIS hanya memiliki 17 item. *Aberrance* yang terdeteksi (1.42%) terbilang hanya sedikit dari keseluruhan jumlah peserta tes. Hal ini dapat disebabkan oleh jumlah item pada EIS yang sedikit.

Pola respon *aberrant* dapat disebabkan oleh banyak variabel, diantaranya adalah motivasi yang rendah, ketelitian, dan beberapa jenis kesalahan lainnya. Pola respon *aberrant* disebabkan oleh perilaku yang menyimpang dari subjek penelitian saat mengerjakan tes, sehingga respon yang diberikan pada tes tidak benar-benar menggambarkan *latent trait* dari subjek, sehingga mempengaruhi dari hasil tes, atau dapat dikatakan hasil pengukuran kurang akurat. Pola respon *aberrant* pada penelitian ini juga

mungkin disebabkan oleh EIS yang tergolong *low stake test*, sehingga beberapa peserta tes cenderung memiliki motivasi dan kinerja yang rendah dalam mengisi EIS.

Berdasarkan uji ANOVA diketahui bahwa  $p = .056$  untuk perbedaan berdasarkan jenis kelamin. Nilai  $p (.056)$  menunjukkan perbedaan yang tidak signifikan antara jenis kelamin peserta tes terhadap  $t_z$  value. Sehingga dapat diketahui bahwa tidak terdapat perbedaan antara peserta tes laki-laki dan perempuan dalam kecenderungan menunjukkan respon yang menyimpang saat pengerjaan tes. Sehingga dapat diketahui bahwa tidak terdapat perbedaan antara peserta tes laki-laki dan perempuan dalam kecenderungan menunjukkan respon yang menyimpang saat pengerjaan tes. Hasil penelitian ini berbeda dengan penelitian terdahulu yang dilakukan oleh Schmitt, Chan, dan Sacco (1999) yang juga menghubungkan *person-fit* dengan jenis kelamin peserta tes. Penelitian Schmitt, Chan, dan Sacco (1999), menemukan bahwa jenis kelamin yang memiliki probabilitas yang lebih tinggi menunjukkan pola respon menyimpang adalah laki-laki. Perbedaan hasil penelitian ini dapat dipengaruhi oleh berbagai variabel, misalnya perbedaan budaya. Tidak terdapatnya perbedaan antara laki-laki dan perempuan juga dapat berkaitan dengan panjang tes. Panjang tes berkaitan dengan validitas tes (Bell & Lumsden, 1980), dimana tes yang terlalu panjang akan menyebabkan peserta tes kelelahan dalam mengerjakan tes sehingga berpengaruh pada skor akhir. Meskipun berdasarkan penelitian terdahulu yang menyatakan bahwa laki-laki cenderung memiliki motivasi



rendah, dan cenderung lebih banyak melakukan perilaku curang dalam dunia pendidikan. Mengingat item pada EIS sedikit (17 item) sehingga terdapat kemungkinan bahwa peserta tes laki-laki tetap dapat berkonsentrasi dan mengerjakan tes dengan sungguh-sungguh.

Selanjutnya, dapat diketahui nilai  $p = .002$  untuk perbedaan  $I_z$  value berdasarkan etnik peserta tes. Nilai  $p$  (.002) menunjukkan terdapatnya perbedaan yang signifikan antar etnik. Hal ini menunjukkan bahwa terdapat etnik tertentu yang lebih memungkinkan dalam menunjukkan pola respon *aberrant* dibanding etnik lainnya.

Penelitian ini menggunakan subjek sebagai peserta tes dengan etnik yang beragam. Beberapa etnik hanya diwakili oleh sedikit orang, sehingga tidak mampu merepresentasikan etnik-etnik terkait. Apabila beberapa etnik yang diwakilkan oleh jumlah peserta tes yang minim diabaikan, maka dapat dilihat bahwa etnik dengan jumlah peserta tes yang cukup besar adalah etnik Jawa dan Tionghoa. Apabila membandingkan nilai *mean* antara kedua etnik tersebut, maka dapat diketahui bahwa peserta tes dari etnik Tionghoa memiliki nilai *mean* yang lebih rendah. Perbandingan nilai *mean* tersebut menunjukkan bahwa peserta tes dari etnik Tionghoa memiliki probabilitas yang lebih tinggi menunjukkan respon *aberrant* dibanding kelompok peserta tes dari etnik Jawa. Guna mengetahui penyebab perbedaan ini, maka dibutuhkan penelitian lanjutan untuk mengkajinya.

Respon *aberrant* sendiri terbagi menjadi 8 macam, yaitu *cheaters*, *lucky guessers*, *flumbers*, *plodders*, *language deficiency*, *random guessers*,

*creatives*, dan *palm beachers*. *Cheaters* merujuk pada peserta tes yang melakukan kecurangan sehingga dapat memperoleh skor yang tinggi. *Lucky guessers* merujuk pada peserta tes yang menebak-nebak jawaban dengan skor tinggi pada tes. *Flumbers* merujuk pada peserta tes yang kebingungan pada awal pengerjaan tes sehingga mendapatkan skor yang rendah. *Plodders* merujuk pada peserta tes yang tidak menyelesaikan tes hingga item terakhir. *Languange deficiency* merujuk pada peserta tes yang tidak memahami bahasa yang digunakan pada tes. *Random guessers* merujuk pada peserta tes yang menebak-nebak jawaban masing-masing item secara acak. *Creatives* merujuk pada peserta tes yang merespon tanpa memperhatikan konten tiap item. *Palm beachers* merujuk pada peserta tes yang salah dalam mengisi lembar jawaban (Meijer, 1996).

Peserta tes yang terdeteksi sebagai *aberrance* bisa saja termasuk ke dalam salah satu dari delapan macam *aberrance* di atas. Namun, pada penelitian ini peserta tes dengan  $I_z$  value ( $<-2.00$ ) tidak dapat dilabel sebagai salah satu dari delapan jenis *aberrance* di atas. Hal ini disebabkan penggunaan statistik  $I_z$  tidak mampu mendeteksi lebih jauh peserta tes *aberrance* berdasarkan jenis *aberrancenya*. Apabila ingin mengetahui pengelompokan *aberrance* secara lebih rinci, maka peneliti perlu untuk melengkapi dan melakukan uji lanjut untuk melakukan analisis, misalnya dengan melakukan wawancara dengan peserta tes (Meijer, 2008), atau menghubungkan  $I_z$  value dengan variabel lainnya (Conjin, 2013).

#### D. Keterbatasan Penelitian

Penelitian ini hanya menggunakan satu statistik *person-fit* sehingga tidak dapat mengetahui perbandingan jumlah *aberrant* yang terdeteksi dengan statistik *person-fit* yang lain. Penggunaan lebih dari satu statistik *person-fit* memungkinkan hasil penelitian dapat menunjukkan perbedaan aplikasi statistik *person-fit* dalam mendeteksi *aberrant*.

Subjek penelitian ini tidak disyaratkan berasal dari etnik tertentu, sehingga beberapa etnik yang ada pada *data set* tidak mampu mewakili atau merepresentasikan  $l_z$  value pada etnik-etnik dari subjek mengingat jumlah subjek pada etnik-etnik tertentu sangat minim. Penentuan etnik tertentu untuk dibandingkan  $l_z$  value antar kelompok akan lebih mampu merepresentasikan respon *aberrant* dari data set.

Skala yang digunakan sebagai alat tes merupakan tes dengan golongan *low stake test*, yang memungkinkan motivasi rendah saat mengerjakannya. Namun, alat tes pada penelitian ini memiliki item yang sedikit. Secara umum, akurasi statistik  $l_z$  berkaitan dengan panjang tes, semakin pendek sebuah tes atau semakin sedikit item pada tes maka akurasi  $l_z$  semakin menurun (Schmitt, Chan, & Sacco., 1999).

## BAB V

### PENUTUP

#### A. Kesimpulan

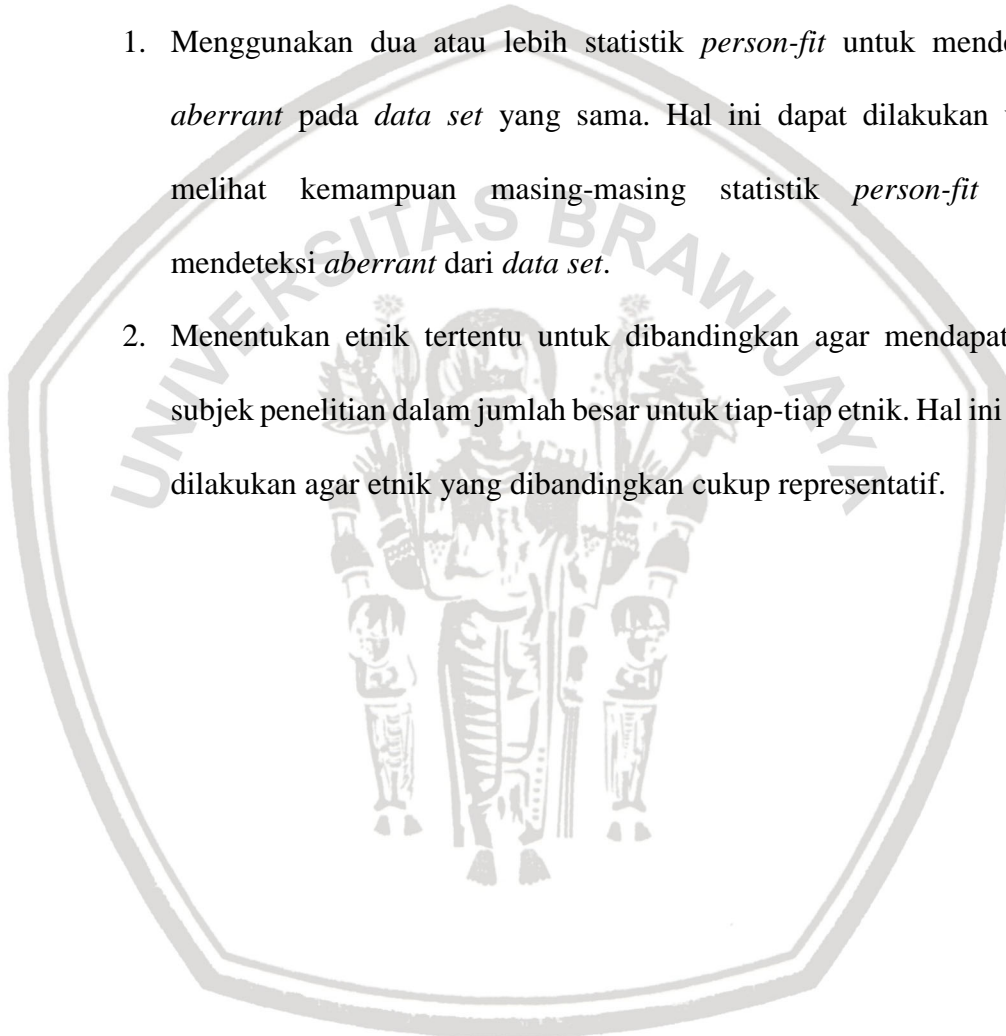
Berdasarkan hasil penelitian mengenai *person-fit* statistik  $l_z$  pada item *polytomous* dan hubungannya dengan jenis kelamin dan etnik, dapat diambil beberapa kesimpulan, antara lain:

1. Hasil penelitian menunjukkan bahwa dari data set dengan 560 subjek, statistik  $l_z$  mendeteksi 8 subjek dengan  $l_z$  value (-2.00). Subjek dengan  $Lz$  value (-2.00) dianggap *aberrance*, dan 552 subjek lainnya dianggap sebagai *non-aberrance*.
2. Tidak terdapat perbedaan antara peserta tes laki-laki dan perempuan.
3. Terdapat perbedaan yang signifikan antara  $l_z$  value dengan etnik peserta tes.
4. Berdasarkan nilai mean dari  $l_z$  value, diketahui bahwa peserta tes etnik Tionghoa memiliki nilai *mean* terendah dibandingkan dengan peserta tes dari etnik Jawa. Hal tersebut menunjukkan bahwa peserta tes dengan etnik Tionghoa memiliki probabilitas yang lebih tinggi dalam memberikan respon yang menyimpang saat mengerjakan tes dibanding etnik Jawa.

## B. Saran

Beberapa usulan saran yang sekiranya dapat dilakukan oleh peneliti selanjutnya jika akan melakukan penelitian dengan tema yang sama, yaitu:

1. Menggunakan dua atau lebih statistik *person-fit* untuk mendeteksi *aberrant* pada *data set* yang sama. Hal ini dapat dilakukan untuk melihat kemampuan masing-masing statistik *person-fit* untuk mendeteksi *aberrant* dari *data set*.
2. Menentukan etnik tertentu untuk dibandingkan agar mendapat data subjek penelitian dalam jumlah besar untuk tiap-tiap etnik. Hal ini dapat dilakukan agar etnik yang dibandingkan cukup representatif.





## DAFTAR PUSTAKA

- Baker, F. (2001). *The basics of item response theory*. United States of America: ERIC Clearinghouse on Assessment and Evaluation.
- Chavez, A.F., & Guido-DiBrito, F. (1999). Racial and ethnic identity and development. *New Direct6ions for Adult and Continuing Education*. doi:10.1002/Ace.8405
- Cohen, R.J. & Swerdlik, M.E. (2005) *Psychological testing and assessment (6th Edition)*. New York: McGraw Hill
- Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.
- DeMars, C., Bashkov, B. M., & Socha, A. B. (2013). The role of gender in test taking motivation under low-stakes condition. *Research and Practice in Assessment*, 8 69-82
- Dragow, F., Levine, M. V., & Williams, E. A. (1985). Appropriateness measurement with polychotomous item response models and standardized indices. *British Journal of Mathematical and Statistical Psychology*, 38, 67–86. doi:1111/j.2044-8317.1985.tb00817.x
- Dragow, F., Levine, M. V., & McLaughlin, M. E. (1991). Appropriateness measurement for some multidimensional test batteries. *Applied Psychological Measurement*, 15, 171–191. doi:10.1177%2F014662169101500207
- Dragow, F., Levine, M. V., & McLaughlin, M. E (1998). A comparison of the person response function and the  $I_z$  person-fit statistic. *Applied Psychological Measurement Journal* Vol. 22. doi:10.1177%2F01466216980221004
- Emons, W. H. M. (2008). Nonparametric person-fit analysis of polytomous item scores. *Applied Psychological Measurement*, 32, 224–247. doi:10.1177%2F0146621607302479
- Honny, J. M., Gadbury-Amyot, C. C., Overman, P. R., Wilkins, K., & Petersen, F. (2010). Academic integrity violations: A national study of dental hygiene students. *Journal of Dental Education*, 74(3), 251-260.

- Horner, M. S. (2005). *Toward an understanding of achievement-related conflicts in women. In M. Ednick, S. Tangi, & L. W. Hoffman (Eds.), Women and achievement. Social and motivational analyses.* John Wiley and Sons.
- Karabatsos, G. (2003) Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Applied Measurement in Education* 16:4, 277-298. doi:10.1207/S15324818AME1604\_2
- Kobayashi, E., & Fukushima, M. (2012). Gender, social bond, and academic cheating in Japan. *Sociological Inquiry*, 82(2), 282-304.
- Levine, M. V., & Rubin, D. B. (1979). Measuring the appropriateness of multiple-choice test scores. *Journal of Educational Statistics*, 4, 269-290. doi:10.2307/1164595
- Meijer, R. R. (1996). Person-fit research: an introduction. *Applied Measurement in Education*, 9(1), 3-8. doi.: 10.1207/s15324818ame0901\_2
- Meijer, R. R. (1997). Person fit and criterion-related validity: An extension of the Schmitt, Cortina, and Whitney study. *Applied Psychological Measurement*, 21, 99-113. doi:10.1177/01466216970212001
- Meijer, R. R., & Sijtsma, K. (2001). Methodology review: Evaluating person fit. *Applied Psychological Measurement*, 25, 107-135. doi:10.1177/01466210122031957
- Ningsih, N. W. (2017). Delayed alternate form reliability skala identitas etnik pada remaja akhir (Desain A - A'). *Skripsi*, Universitas Brawijaya, Psikologi, Malang.
- Paramita, W. (2014). Perbedaan identitas etnis pada remaja etnis jawa dan etnis tionghoa. *Skripsi*, Universitas Brawijaya, Psikologi, Malang.
- Phinney, J. (1992). The Multigroup ethnic Identity Measure: A New Scale for Use with Diverse Groups. *Journal of Adolescent Research*, 7(156), 156-176. doi:10.1177/074355489272003
- Phinney, J., & Ong, A. D. (2007). Conceptualization and measurement of ethnic identity: Current status and future directions. *Journal of Counseling Psychology*, 54(3), 271- 281. doi:10.1037/0022-0167.54.3.271
- Salleh, M. I. M., Hamid, H. A., Alias, N. R., Ismail, M. N., & Yusoff, Z. (2011). *The influence of gender and age on the undergraduate's academic dishonesty behaviours. In Sociality and Economics Development (pp. 593-597).* Kuala Lumpur, Malaysia: Book series 10, International Proceedings of Economics Development and Research

- Santrock,, J. W.(2009). *Perkembangan Anak. Edisi 11*. Jakarta. Erlangga
- Saulsbury, M. D., Brown, U. J., Heyliger, S. O., & Beale, R. L. (2011). Effect of dispositional traits on Pharmacy students' attitude towards cheating. *Journal of Pharmaceutical Education*, 75(4), 1-8.
- Schmitt, Chan, & Sacco. (1999). Correlates of person fit and effect of person fit on test validity. *Applied Psychological Measurement*, Vol. 23 No. 1, March 1999, 41–53. doi:10.1177%2F01466219922031176
- Smith, R. M. (1986). Person fit in the Rasch model. *Educational and Psychological Measurement*, 46, 359–372. doi:10.1177%2F001316448604600210
- Tatsuoka, K. K. (1984). Caution indices based on item response theory. *Psychometrika*, 49, 95–110. doi:10.1007/BF02294208
- Umaña-Taylor, A., Yazedjian, A., & Bámaca-Gómez, M. (2004). Developing the ethnic identity scale using eriksonian and social identity perspectives. *identity: An International Journal of theory and Research*, 4(1), 9-38. doi:10.1207/S1532706XID0401\_2
- van Krimpen-Stoop, E. M. L. A., & Meijer, R. R. (1999). The null distribution of person- fit statistics for conventional and adaptive tests. *Applied Psychological Measurement*, 23, 327–345. doi:10.1177%2F01466219922031446
- Wainer, H. (1993). Measurement problems. *Journal of Educational Measurement*, 30, 1–21. doi:10.1111/j.1745-3984.1993.tb00419.x
- Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, 10(1), 1–17. doi:10.1207/s15326977ea1001\_1
- Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago, IL: Mesa Press.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis [Computer Software]*. Chicago, IL: Mesa Press.